
Minimax Error of Interpolation and Optimal Design of Experiments for Variable Fidelity Data

A.Zaytsev

a.zaytsev@skoltech.ru

E.Burnaev

e.burnaev@skoltech.ru

Skolkovo Institute of Science and Technology (Skoltech),
Insitute for Information Transmission Problems (IITP RAS)

Abstract

Engineering problems often involve data sources of variable fidelity with different costs of obtaining an observation. In particular, one can use both a cheap low fidelity function (e.g. a computational experiment with a CFD code) and an expensive high fidelity function (e.g. a wind tunnel experiment) to generate a data sample in order to construct a regression model of a high fidelity function. The key question in this setting is how the sizes of the high and low fidelity data samples should be selected in order to stay within a given computational budget and maximize accuracy of the regression model prior to committing resources on data acquisition.

In this paper we obtain minimax interpolation errors for single and variable fidelity scenarios for a multivariate Gaussian process regression. Evaluation of the minimax errors allows us to identify cases when the variable fidelity data provides better interpolation accuracy than the exclusively high fidelity data for the same computational budget.

These results allow us to calculate the optimal shares of variable fidelity data samples under the given computational budget constraint. Real and synthetic data experiments suggest that using the obtained optimal shares often outperforms natural heuristics in terms of the regression accuracy.

1 INTRODUCTION

In some cases sample data for regression modeling has variable fidelity: some data comes from a high fidelity source, some – from a low fidelity source [12]. While there are many approaches to handle variable fidelity data including transfer learning [24] and space mapping [3] techniques, engineers often use cokriging approach [17] based on the Gaussian process framework [37, 28]. Numerous applications of cokriging include geostatistics [35], aerospace [14], and engineering [19]. In this paper we also consider this approach for modeling data, obtained from high and low fidelity data sources.

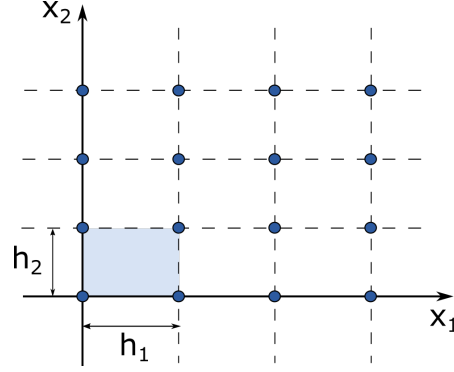
The interest in accuracy of Gaussian process models for single fidelity data dates back to Wiener and Kolmogorov [34, 18]. They obtained an error at a specified point in the univariate case. Further progress in refining this estimate is available in the book by Stein [30], inspired by Ibragimov results [15]. Recent results expand this setting by considering a more general interpolation error, equal to the integral of the squared difference between the true function and an its interpolation over the domain of interest, see [32] for finite sample results in the multivariate case, and [13] for results about the minimax error of interpolation over an infinite regular univariate sample. References [39, 31, 6], to name a few, report similar results.

While in case of single fidelity data results are quite well established, there is only one paper [38], to our knowledge, that investigates the interpolation error for the variable fidelity data case from a theoretical point of view. For a squared exponential covariance function and a squared error at a single point, authors identify cases when regression modeling based on variable fidelity data is superior to using only the high fidelity data. Other papers dealing with variable fidelity regression modeling, [33, 5, 25], focus on statistical properties of regression parameters estimates, but provide little insight into understanding how and why the variable fidelity modeling works.

Due to current apparent scarcity of theoretical foundations practitioners usually adopt heuristic rules in determining sizes of data samples of different fidelity and quantify when to use the variable fidelity data [2, 29, 21]; or they use adaptive design of experiments approaches and surrogate based optimization directly, see [27, 16, 8, 22] and references therein.

The main contributions of this paper are the following:

- **Minimax interpolation error for the multivariate case.** We start with obtaining the interpolation error for the Gaussian process regression with a known covariance function. Then we derive the minimax interpolation error for functions from a general smoothness class in the multivariate case. This error is a nontrivial generalization of the univariate results obtained in [13].
- **The optimal ratio of sizes of variable fidelity data samples.** We obtain the interpolation error for the specified covariance function in the variable fidelity case, and then derive the minimax interpolation error in the general additive setting (cokriging) [17]. With the derived minimax interpolation error we identify when and to which extent the variable fidelity regression modeling is beneficial compared to the regression modeling using only a high fidelity data under the same computational budget. We calculate the optimal ratio of sizes of variable fidelity data samples given the budget constraint. There is a certain gap between the theoretical setup we consider and the real world: we consider a setting that uses an infinite grid as a design of experiments and requires knowledge of relative complexities of high and low fidelity functions to calculate the optimal ratio of sample sizes. Nevertheless these theoretical results are sufficient to provide justification for the corresponding applied algorithm we develop.
- **The technique to select the ratio of sizes of variable fidelity data samples.** We elaborate on a method to choose the ratio inspired by our theoretical investigations. While the existing approaches usually work in adaptive design of experiments setting and pick points using sufficiently accurate regression models constructed beforehand [27], we offer a method to select sizes of high and low fidelity data samples to fit into a given computational budget and maximize accuracy of a resulting regression model prior to spending any


 Figure 1: Design D_H for $d = 2$.

significant resources on data generation. Our estimate depends only on the computational cost of variable fidelity data generation and on a correlation between high and low fidelity functions. We investigate the applicability of the proposed technique by comparing it to a number of natural baselines on synthetic and real datasets.

We provide proofs of all theorems in appendices.

2 Minimax Interpolation Error for Gaussian Process Regression

In case of Gaussian process regression modeling there is a gap between theoretically tractable problems and practice. Namely, since the main tool for theoretical investigation is the Fourier transform, it is a common approach to consider the design of experiments based on an infinite grid [13, 30], though in many cases the theoretical results are transferable to practical solutions. Thus, in this section we consider a design of experiments, belonging to some infinite grid, and later in the experimental section we show that our conclusions remain valid under finite sample random designs.

2.1 Interpolation Error

Let $f(\mathbf{x})$ be a stationary Gaussian process on \mathbb{R}^d with a covariance function $R(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}_0 + \mathbf{x}) - \mathbb{E}f(\mathbf{x}_0 + \mathbf{x}))(f(\mathbf{x}_0) - \mathbb{E}f(\mathbf{x}_0))$ and a spectral density $F(\boldsymbol{\omega})$

$$F(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} e^{2\pi i \boldsymbol{\omega}^T \mathbf{x}} R(\mathbf{x}) d\mathbf{x}.$$

Suppose that we know values of realizations of $f(\cdot)$ at the infinite rectangular grid $D_H = \{\mathbf{x}_{\mathbf{k}} : \mathbf{x}_{\mathbf{k}} = H\mathbf{k}, \mathbf{k} \in \mathbb{Z}^d\}$, where H is a diagonal matrix with elements h_1, \dots, h_d . An example of such design in the case of the input dimension $d = 2$ is provided in Figure 1.

We measure the interpolation error over the domain of interest $\Omega_H = [0, h_1] \times \dots \times [0, h_d]$ as follows:

$$\sigma_H^2(\tilde{f}, F) = \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E} [\tilde{f}(\mathbf{x}) - f(\mathbf{x})]^2 d\mathbf{x}, \quad (1)$$

where $\mu(\Omega_H) = \prod_{i=1}^d h_i$ is the Lebesgue measure of Ω_H , and $\tilde{f}(\mathbf{x})$ is the interpolation of $f(\mathbf{x})$. Here we consider $\tilde{f}(\mathbf{x})$ having the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}_{\mathbf{k}}), \quad (2)$$

where $K(\cdot)$ is a symmetric kernel.

Theorem 1. *The error of interpolation with $\tilde{f}(\mathbf{x})$ from (2), based on observations at points from D_H of a stationary Gaussian process $f(\mathbf{x})$ with spectral density $F(\boldsymbol{\omega})$, is equal to*

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left(1 - \hat{K}(\boldsymbol{\omega})\right)^2 + \sum_{\mathbf{x} \in D_{H^{-1}} \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] d\boldsymbol{\omega},$$

where $\hat{K}(\boldsymbol{\omega})$ is the Fourier transform of $K(\boldsymbol{\omega})$. Furthermore, the optimal $\hat{K}(\boldsymbol{\omega})$, minimizing the interpolation error, has the form

$$\hat{K}(\boldsymbol{\omega}) = \frac{F(\boldsymbol{\omega})}{\sum_{\mathbf{x} \in D_{H^{-1}}} F(\boldsymbol{\omega} + \mathbf{x})}.$$

Remark 1. *The function $\tilde{f}(\mathbf{x})$ that minimizes the squared error $\mathbb{E}(\tilde{f}(\mathbf{x}) - f(\mathbf{x}))^2$ has the form (2), where $K(\cdot)$ is a symmetric kernel. This motivates us to use $\tilde{f}(\mathbf{x})$ from (2) for interpolation.*

Remark 2. *It is easy to see that for $\tilde{f}(\mathbf{x})$ from (2) it holds that*

$$\sigma_H^2(\tilde{f}, F) = \sigma_{S_H}^2(\tilde{f}, F),$$

where $S = \text{diag}(s_1, \dots, s_d)$, with $s_i \in \mathbb{Z}^+, i = 1, \dots, d$.

Using Theorem 1 one can estimate interpolation errors for various covariance functions. For example,

Corollary 1. *For a Gaussian process on \mathbb{R} with exponential spectral density $F_\theta(\omega) = \frac{\theta}{\theta^2 + \omega^2}$ the interpolation error (1) has the form:*

$$\sigma_h^2(\tilde{f}, F_\theta) \approx \frac{2}{3}\pi^2\theta h + O((\theta h)^2), \quad \theta h \rightarrow 0.$$

Corollary 2. *For a Gaussian process on \mathbb{R} with squared exponential spectral density $F_\theta(\omega) = \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right)$ the interpolation error (1) is bounded by:*

$$\frac{4}{3}h\sqrt{\theta} \exp\left(-\frac{1}{8h^2\theta}\right) \leq \sigma_h^2(\tilde{f}, F_\theta) \leq 7h\sqrt{\theta} \exp\left(-\frac{1}{8h^2\theta}\right), \quad \theta h^2 \rightarrow 0.$$

2.2 Minimax Interpolation Error

For many covariance functions direct evaluation of the interpolation error can be technically cumbersome, especially for the case $d > 1$. Furthermore, in many cases the true covariance function is not known exactly, and calculating the interpolation error in such misspecified cases is even a harder task.

Instead we consider a minimax interpolation error that provides an answer in the worst case scenario. We define a set $\mathcal{F}(L, \boldsymbol{\lambda})$ of spectral densities $F(\boldsymbol{\omega})$ for a given $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$ and $L > 0$ as

$$\mathcal{F}(L, \boldsymbol{\lambda}) = \left\{ F : \mathbb{E} \sum_{i=1}^d \lambda_i^2 \left(\frac{\partial f_F(\mathbf{x})}{\partial x_i} \right)^2 \leq L, \mathbf{x} \in \mathbb{R}^d \right\}, \quad (3)$$

where $f(\mathbf{x}) = f_F(\mathbf{x})$ is a Gaussian process with the spectral density $F(\boldsymbol{\omega})$ at the point $\mathbf{x} \in \mathbb{R}^d$. Sample realizations of Gaussian processes for different L in the case of $d = 1$ and the Matérn covariance function [28] are shown in Figure 2.

The minimax interpolation error is defined as follows:

$$R^H(L, \boldsymbol{\lambda}) = \inf_{\tilde{f}} \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \sigma_H^2(\tilde{f}, F).$$

This error describes how large the interpolation error is for the worst case scenario. The following theorem holds:

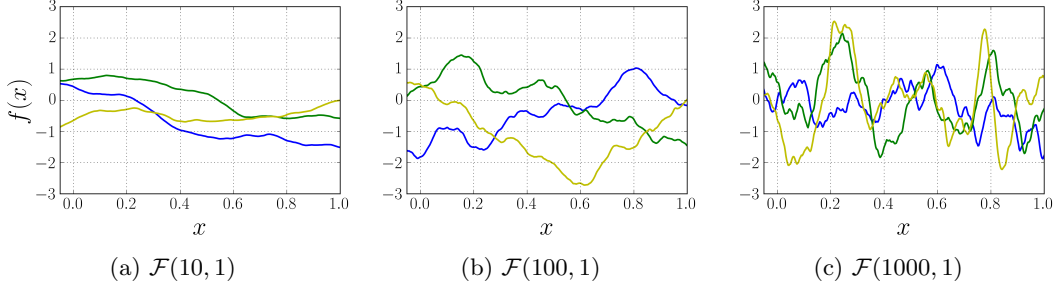


Figure 2: Realizations of Gaussian processes with the Matérn covariance function $R(x) = (1 + \sqrt{3}\theta|x|) \exp(-\sqrt{3}\theta|x|)$ ($\nu = \frac{3}{2}$) for different values of L in $\mathcal{F}(L, 1)$ and $d = 1$.

Theorem 2. For a Gaussian process $f(\mathbf{x})$, defined on \mathbb{R}^d and evaluated on the design D_H , with the spectral density belonging to the set $\mathcal{F}(L, \boldsymbol{\lambda})$, the minimax interpolation error has the form

$$R^H(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i} \right)^2.$$

Moreover, the minimax optimal interpolation $\tilde{f}(\mathbf{x})$ has the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}'),$$

where $K(\mathbf{x})$ is a symmetric kernel with the Fourier transform $\hat{K}(\boldsymbol{\omega})$ defined as

$$\hat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \sqrt{\sum_{i=1}^d \omega_i^2 \cdot h_i^2} & \text{if } \sum_{i=1}^d \omega_i^2 \cdot h_i^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note, that we can minimize the minimax interpolation error w.r.t. the diagonal matrix H in such a way as to keep fixed the average number of points belonging to a unit hypercube: $\prod_{i=1}^d \frac{1}{h_i} = n$. The corresponding optimal matrix $H^* = \text{diag}(h_1^*, \dots, h_d^*)$ has the form:

$$h_i^* = \sqrt[d]{\frac{n \lambda_i^d}{\prod_{j=1}^d \lambda_j}}.$$

The minimal minimax interpolation error is then equal to $R^{H^*}(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \sqrt[d]{\frac{n}{\prod_{i=1}^d \lambda_i}}$.

3 Minimax Interpolation Error for a Variable Fidelity Model

3.1 Variable Fidelity Data Model

Suppose that the true function is modelled as

$$u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x}), \quad (4)$$

where ρ is a fixed constant, and $f(\mathbf{x})$ and $g(\mathbf{x})$ are stationary independent Gaussian processes, defined on \mathbb{R}^d . This is the state-of-the-art cokriging approach used to model a variable fidelity data [17].

We refer to a realization of $u(\mathbf{x})$ as a high fidelity function, and to a realization of $f(\mathbf{x})$ as a low fidelity function. Therefore $g(\mathbf{x})$ is a correction of $f(\mathbf{x})$ that appears due to a low fidelity nature of $f(\mathbf{x})$. The parameter ρ provides information on a strength of the relation between $f(\mathbf{x})$ and $u(\mathbf{x})$.

We observe values of $u(\mathbf{x})$ and $f(\mathbf{x})$ and we want to construct an interpolation $\tilde{u}(\mathbf{x})$ of the high fidelity function $u(\mathbf{x})$ on the basis of these variable fidelity observations.

3.2 Interpolation Error

It is natural to assume that we observe the cheap low fidelity function $f(\mathbf{x})$ on denser grid than the expensive high fidelity function $u(\mathbf{x})$. We observe $u(\mathbf{x})$ at points from $D_u = D_H$, and $f(\mathbf{x})$ at points from $D_f = D_{\frac{H}{m}}$ with $m \in \mathbb{Z}^+$.

Using these observations we attempt to interpolate $u(\mathbf{x})$ within the hypercube $[0, h_1] \times \dots \times [0, h_d]$. The function $\tilde{u}(\mathbf{x})$ minimizes the interpolation error $\sigma_{H,m}^2(\tilde{u}, F, G, \rho)$ for observations of $u(\mathbf{x})$ over D_H and observations of $f(\mathbf{x})$ over $D_{\frac{H}{m}}$:

$$\sigma_{H,m}^2(\tilde{u}, F, G, \rho) = \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E} [\tilde{u}(\mathbf{x}) - u(\mathbf{x})]^2 d\mathbf{x}. \quad (5)$$

Theorem 3. *The minimum of interpolation error (5) of the variable fidelity data model $u(\mathbf{x})$ from (4), based on observations of $u(\mathbf{x})$ at points from D_H and observations of $f(\mathbf{x})$ at points from $D_{\frac{H}{m}}$, has the form:*

$$\sigma_{H,m}^2(\tilde{u}, F, G, \rho) = \sigma_H^2(\tilde{g}, G) + \rho^2 \sigma_{\frac{H}{m}}^2(\tilde{f}, F), \quad (6)$$

where $\tilde{g}(\mathbf{x})$ and $\tilde{f}(\mathbf{x})$ minimize $\sigma_H^2(\tilde{g}, G)$ and $\sigma_{\frac{H}{m}}^2(\tilde{f}, F)$ respectively.

3.3 Minimax Interpolation Error

We obtain the minimax interpolation error for the variable fidelity case in the manner similar to the single fidelity case. Let us assume that the true spectral densities of the processes $f(\cdot)$ and $g(\cdot)$ are unknown, but sufficiently smooth, i.e. they belong to classes $\mathcal{F}(L_f) = \mathcal{F}(L_f, \mathbf{1})$ and $\mathcal{F}(L_g) = \mathcal{F}(L_g, \mathbf{1})$ respectively. Here for clarity of the presentation we limit ourselves to the case $\boldsymbol{\lambda} = \mathbf{1} \in \mathbb{R}^d$ and $H = h\mathbf{I}$ for some $h > 0$, where $\mathbf{I} \in \mathbb{R}^{d \times d}$. In fact, results below hold in a more general setting, described in section 2 and defined by general values of $\boldsymbol{\lambda} \in \mathbb{R}^d$ and H . However, this additional sophistication blurs the main conclusions and provides little additional insight.

The goal is to obtain the minimax interpolation error for $u(\mathbf{x})$. In particular we want to get the minimax interpolation error for the variable fidelity data

$$R^{h,m}(L_f, L_g) = \inf_{f,g} \sup_{\substack{F \in \mathcal{F}(L_f), \\ G \in \mathcal{F}(L_g)}} \sigma_{h\mathbf{I},m}^2(\tilde{u}, F, G, \rho). \quad (7)$$

Theorem 4. *Minimax interpolation error (7) of model (4), based on observations of $u(\mathbf{x})$ at points from D_H and observations of $f(\mathbf{x})$ at points from $D_{\frac{H}{m}}$, has the form*

$$R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{h}{m\pi} \right)^2 + \frac{L_g}{2} \left(\frac{h}{\pi} \right)^2. \quad (8)$$

4 Optimal Ratio of sizes of Variable Fidelity Data Samples

Obtained results allow us to get the optimal ratio m of sizes of variable fidelity data samples. We consider the following setting: one evaluation of $u(\mathbf{x})$ costs c , whereas one evaluation of $f(\mathbf{x})$ is 1; the total evaluation cost is equal to the number of points in a unit hypercube $\frac{1}{h^d}$ multiplied by the evaluation price; and the computational budget is set to Λ .

For such setup the total budget is equal to $c \frac{1}{h^d} + \delta \frac{1}{h^d}$, where $\delta = m^d$ is the ratio of sizes of variable fidelity data samples.

Using Theorem 4 we prove

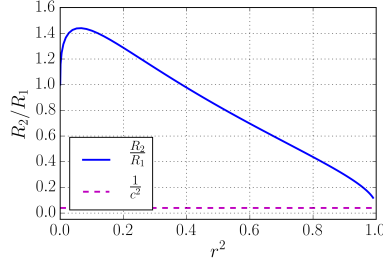


Figure 3: Dependence of the ratio $\frac{R_2}{R_1}$ of the minimax interpolation errors on the correlation coefficient r for $L_f = 3, L_g = 1, c = 5, d = 1$.

Theorem 5. *The minimum of the minimax interpolation error (8) given the computational budget Λ has the form*

$$\min_{\substack{h, \delta: \\ \Lambda h^d = c + \delta}} R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{c + \delta^*}{\pi \Lambda \delta^*} \right)^{\frac{2}{d}} + \frac{L_g}{2} \left(\frac{c + \delta^*}{\pi \Lambda} \right)^{\frac{2}{d}},$$

and the optimal ratio is $\delta^* = \left(\frac{L_f}{L_g} c \rho^2 \right)^{\frac{d}{d+2}}$.

The optimal ratio δ^* depends on the relative cost c of the high fidelity function evaluation, the coefficient ρ and the smoothnesses L_f and L_g of $f(\mathbf{x})$ and $g(\mathbf{x})$ respectively.

If for the interpolation we use evaluations of $u(\mathbf{x})$ exclusively, then we get the following minimax interpolation error given the budget Λ :

$$\min_{h: \Lambda h^d = c} R^h(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{c}{\pi \Lambda} \right)^{\frac{2}{d}} + \frac{L_g}{2} \left(\frac{c}{\pi \Lambda} \right)^{\frac{2}{d}}.$$

Note, that we can get similar results for a specific covariance function using Theorem 3 and Corollaries 1 and 2.

4.1 Comparison of Minimax Interpolation Errors Under Different Scenarios

Let us now investigate under what conditions and to what extent the usage of the variable fidelity data can decrease the interpolation error compared to using single fidelity data within the same computational budget. We denote by $R_2 = R^{h, \delta^*}(L_f, L_g, \rho)$ the minimax interpolation error, obtained when using the variable fidelity data, and by $R_1 = R^h(L_f, L_g, \rho)$ the minimax interpolation error, obtained when using only the high fidelity data. The ratio $\frac{R_2}{R_1}$ characterizes benefits of the variable fidelity data over single fidelity data: $\frac{R_2}{R_1} \geq 1$ means there is no advantage to using the variable fidelity data, while $\frac{R_2}{R_1} < 1$ implies that the variable fidelity data improves the accuracy of the interpolation.

The ratio $\frac{R_2}{R_1}$ has the form:

$$\frac{R_2}{R_1} = \frac{\left(1 + \left(\frac{L_f^d \rho^{2d}}{L_g^d c^2} \right)^{\frac{1}{d+2}} \right)^{\frac{d+2}{d}}}{1 + \rho^2 \frac{L_f}{L_g}}.$$

If we put $V_f = \mathbb{E}f^2(\mathbf{x})$ and $V_g = \mathbb{E}g^2(\mathbf{x})$, then the correlation coefficient r between $u(\mathbf{x})$ and

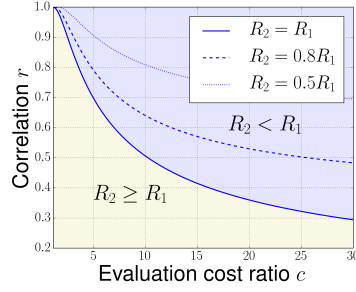


Figure 4: Curves $R_2 = kR_1$ for $L_f = 2$, $L_g = 1$, $d = 1$.

$f(\mathbf{x})$ is $r = \frac{1}{\sqrt{1 + \frac{V_g}{V_f} \frac{1}{\rho^2}}}$. Thus for $r \rightarrow 0$ or $r \rightarrow 1$ it holds that

$$\begin{aligned} r \rightarrow 0 : \frac{R_2}{R_1} &\approx 1 + \frac{d+2}{d} \left(\frac{L_f V_f}{L_g V_g} \right)^{\frac{d}{d+2}} \frac{r^{\frac{2d}{d+2}}}{c^{\frac{2}{d+2}}}, \\ r \rightarrow 1 : \frac{R_2}{R_1} &\approx \frac{1}{c^{\frac{2}{d}}} + \frac{2+d}{d} \left(\frac{L_g V_f}{L_f V_g} \right)^{\frac{d}{d+2}} \frac{(1-r^2)^{\frac{d}{d+2}}}{c^{\frac{4}{d(d+2)}}}. \end{aligned}$$

If $r \rightarrow 0$ then the variable fidelity data is unable to improve the accuracy of the interpolation, while when $r \rightarrow 1$ the ratio $\frac{R_2}{R_1}$ approaches $\frac{1}{c^{\frac{2}{d}}}$, where usually $c \gg 1$. The speed of convergence improves as c increases and $\frac{L_g}{L_f}$ decreases, which means that if $g(\mathbf{x})$ is smoother than $f(\mathbf{x})$, then the variable fidelity data improves accuracy of interpolation additionally.

In figure 3 we show how the ratio $\frac{R_2}{R_1}$ depends on $r = r(\rho)$ in case of $d = 1$. For small r it holds that $R_2 > R_1$ no matter how large our computational budget is, while for high enough r the value of $\frac{R_2}{R_1}$ tends to $\frac{1}{c^2}$, $c \gg 1$.

Figure 4 depicts the smallest values of r in the case of $d = 1$, which for the fixed $c > 1$ provides $R_2 \leq R_1$.

For $d > 1$ and $c \gg 1$ the minimal value of r that provides $R_2 \leq R_1$ is of the order $\frac{1}{\sqrt{c}}$:

$$r \approx \frac{1}{\sqrt{c}} \left(\frac{V_f}{V_g} \right)^{\frac{1}{2}} \left(\frac{L_g}{L_f} \right)^{\frac{1}{2}}.$$

4.2 Optimal Ratio of Sample Sizes for Variable Fidelity Data

If we know the true covariance function it is easy to estimate the parameters L_f and L_g with the second derivatives of the covariance function $\frac{\partial^2 R(\mathbf{x})}{\partial x_i \partial x_j}$ at the point $\mathbf{x} = \mathbf{0}$. However, in the small sample case it is difficult to estimate the parameters of the covariance function [37] or the sum of partial derivatives [20] accurately.

Therefore, assuming $L_f = L_g$ and using Theorem 4, we propose **Technique 1**, that can be used to estimate the optimal ratio of sample sizes δ^* and produce a design of experiments for the case of variable fidelity data. The advantage of the proposed technique is that it can be used even for a variable fidelity modeling approach different from the Gaussian process regression framework; and it requires little prior knowledge about the dependence structure between the high and the low fidelity functions, in particular, we only have to estimate the correlation coefficient r .

5 Experiments

We evaluate the performance of the proposed algorithm for estimation of the optimal ratio of sample sizes for variable fidelity data in two steps: we start with synthetic data generated

Technique 1 Generation of designs of experiments D_f and D_u for evaluations of the low fidelity function and the high fidelity function respectively.

Input: Correlation r between the variable fidelity observations, budget Λ , cost c of one high fidelity function evaluation (the cost of evaluating the low fidelity function is fixed at 1)

- 1: $\rho^2 \leftarrow 1/(1 - \frac{1}{r^2})$
- 2: $\delta^* \leftarrow (c\rho^2)^{\frac{d}{d+2}}$
- 3: $n_f \leftarrow \frac{\Lambda\delta^*}{c+\delta^*}, n_u \leftarrow \frac{\Lambda}{c+\delta^*}$
- 4: Generate random nested designs of experiments $D_f, D_u, D_u \subseteq D_f$, with $|D_f| = n_f, |D_u| = n_u$.
- 5: **return** D_f, D_u

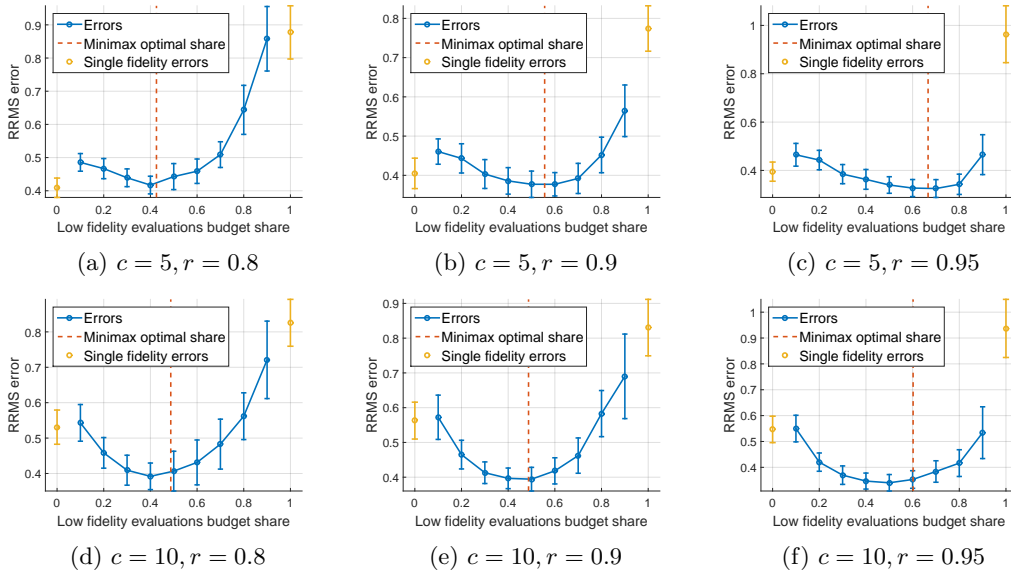


Figure 5: Synthetic data. Dependence of the RRMS error on the share of the budget, allocated for the low fidelity function evaluations. We consider the case of $d = 3$ and different correlations r between the high and low fidelity functions. **Yellow** points correspond to the case when we use either exclusively high or exclusively low fidelity data. The results are averaged across 20 runs.

as realizations of Gaussian processes, and then consider data that originates from engineering applications.

We use the Matérn covariance function $R_{\theta}(\mathbf{x} - \mathbf{x}')$ with $\nu = \frac{3}{2}$ that provides differentiable realizations of Gaussian processes [28]:

$$R_{\theta}(\mathbf{x} - \mathbf{x}') = (1 + \sqrt{3}d_{\theta}(\mathbf{x} - \mathbf{x}')) \exp(-\sqrt{3}d_{\theta}(\mathbf{x} - \mathbf{x}')),$$

where $d_{\theta}(\mathbf{x} - \mathbf{x}') = \sqrt{\sum_{i=1}^d \theta_i(x_i - x'_i)^2}$. To construct a Gaussian process regression model we use Bayesian estimates of the covariance function parameters [9] obtained in a way similar to [10], as open source software alternatives require manual tuning for each particular problem to provide a reasonable comparison [23].

To assess model accuracy we use the Relative Root Mean Squared Error (RRMS) estimated using a dedicated test sample in case of a synthetic data and the cross-validation procedure in case of a real data. For a model $\tilde{u}(\mathbf{x})$ and a test sample $S_* = \{\mathbf{x}_i^*, u_i^* = u(\mathbf{x}_i^*)\}_{i=1}^{n_t}$ the RRMS error is given by $\text{RRMS} = \sqrt{\frac{\sum_{i=1}^{n_t} (u_i^* - \tilde{u}(\mathbf{x}_i^*))^2}{\sum_{i=1}^{n_t} (u_i^* - \bar{u})^2}}$, where $\bar{u} = \frac{1}{n_t} \sum_{i=1}^{n_t} u_i^*$.

Baseline Technique	n_u	n_f
High	Λ/c	0
EqualSize	$\Lambda/(c+1)$	$\Lambda/(c+1)$
EqualBudget	$\Lambda/(2c)$	$\Lambda/2$
Low	0	Λ
MinMinimax	$\Lambda/(c+\delta^*)$	$\delta^*\Lambda/(c+\delta^*)$

Table 1: Sizes of the high fidelity sample (n_u) and the low fidelity sample (n_f) in case of the computational budget Λ .

Data and scripts, used to run the experiments, are at gitlab.com/JohnDoe1989/VariableFidelityData.

5.1 Synthetic Data Experiments

In this section we generate synthetic data as a realization of a Gaussian process with a specified covariance function. We follow the model $u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x})$, with nested designs, i.e. $D_u \subseteq D_f$, and design points picked uniformly at random from $[0, 1]^d$. The total computational budget is 300, and the cost of evaluating $u(\mathbf{x})$ is either $c = 5$ or $c = 10$. Since the exact values of ρ and r are known, we use them in our experiments to calculate δ^* .

Figures 5 depict the dependence of the RRMS error on the proportion of the computational budget allocated for the low fidelity function evaluations. It can be seen that our estimate of the optimal ratio δ^* is close to the true optimal ratio despite the fact that estimates of the unknown parameters of the Gaussian Process regression model were used, and the design of experiments was not a grid.

5.2 Baseline Techniques

We compare our technique for estimation of the optimal ratio of sample sizes, which we call **MinMinimax**, to four different baseline heuristics:

- **High** — only the high fidelity data is used,
- **Low** — we use only the low fidelity data,
- **EqualSize** — the sizes of low and high fidelity data samples are equal,
- **EqualBudget** — the budget is devoted equally to low and high fidelity function evaluations.

Relative sizes of samples for these techniques are given in Table 1.

5.3 Real Data Experiments

We consider the following real data problems. The first three of them (**Euler**, **Airfoil** [4], **MachAngle**) are concerned with calculation of lift and drag coefficients of an airfoil depending on flight conditions and airfoil geometry. To evaluate these outputs we use different solvers for the high and the low fidelity data sources. The next two problems (**Press** [10], **Disk** [36]) investigate dependence of maximum stress and maximum displacement on geometry of the equipment considered. Although three data fidelities are available in the **Press** problem, in each experiment we use only two. The last two problems ([16], **SVM**, **Supernova**) are related to modeling dependence of the goodness-of-fit on model parameters. Details on the problems are provided in the supplementary materials. Input dimensions for these problems are listed in Table 2.

The budget Λ is equal to 300 for all problems except **Euler**, as in this problem the sample size is small. For comparison consistency the cost ratio c is kept at 5 for all given problems. If the

Problem	Output number	Input dimension	High	EqualSize	EqualBudget	MinMinimax	Low
Euler	1	11	0.7674	0.8925	0.8462	0.7420	0.9139
Euler	2	11	0.0668	0.0778	0.2699	0.3803	0.3974
Airfoil	1	6	0.5462	0.5946	0.5390	0.5221	0.4852
Airfoil	2	6	0.1200	0.1422	0.1301	0.1381	0.2962
MachAngle	1	2	0.0886	0.1064	0.1951	0.1951	0.4052
MachAngle	2	2	0.0938	0.1148	0.1714	0.1796	0.3651
Press12	1	6	0.5599	0.6019	0.3580	0.2779	0.2843
Press12	2	6	0.4433	0.4918	0.2715	0.1768	0.1768
Press13	1	6	0.5596	0.5759	0.3861	0.3481	0.5435
Press13	2	6	0.4492	0.4852	0.2782	0.1798	0.1798
Disk	1	6	0.2999	0.3400	0.1922	0.1934	0.1638
Disk	2	6	0.4460	0.4570	0.2998	0.2998	0.2723
SVM	1	2	0.1487	0.1492	0.1849	0.1642	0.6081
Supernova	1	3	0.0395	0.0484	0.0180	0.0575	0.0575

Table 2: RRMS errors averaged over 20 runs of the cross-validation procedure for the problems with real data.

MinMinimax technique returns the sample size $n_u < 1$, then only the low fidelity data is used. For the **MinMinimax** technique we use the correlation coefficient r estimated using the whole available data sample. In addition, to keep the comparison meaningful we normalize all the data before constructing regression models to get variables with zero mean and unit variance.

Euler. Eleven input variables parametrize geometry of an airfoil. We use two different solvers to obtain high and low fidelity values. We use an Euler equations solver to generate the high fidelity data and a full potential equations solver to generate the low fidelity data.

Airfoil. The geometry of an airfoil and the flight regime (the speed and the angle of attack) are described by 52 input variables. We employ a dimension reduction procedure similar to the PCA, and model the dependence on six input factors [4]. We use two different solvers to obtain high and low fidelity values.

MachAngle. Two input variables are the Mach number and the angle of attack for a specific airfoil. We use two different solvers to obtain high and low fidelity values. Low fidelity solver provides almost linear dependence.

Press. We model the maximum stress and the maximum displacement for a C-shaped press [10]. Six input variables describe the geometry of the press, and the fidelity of output depends on a mesh size. We generate three different data samples that correspond to high, moderate and low fidelity outputs. We refer to the case when we model the high fidelity output by $u(\mathbf{x})$ and the moderate fidelity output by $f(\mathbf{x})$ as **Press12**, and the case when we model the high fidelity output by $u(\mathbf{x})$ and the low fidelity output by $f(\mathbf{x})$ as **Press13**.

Disk. We model the maximum stress and the maximum displacement of a rotating disk in an engine [36]. Six input variables describe the geometry of the disk. We use two different solvers to obtain high and low fidelity values.

SVM. We model the dependence of the SVM classifier accuracy from the **sklearn** [26] on the kernel bandwidth and the margin coefficient for the “MAGIC Gamma Telescope” dataset [16]. We have two input variables. As a measure of accuracy we use the area under the ROC curve as suggested by the authors of the dataset [7]. To generate the low fidelity dataset we estimate the accuracy of the classifier constructed using 500 training points, and to generate the high fidelity dataset we estimate the accuracy of the classifier constructed using 2000 training points.

Supernova. We model the dependency of the likelihood of the supernova redshift data on the three fundamental physical constants, similarly to [16]. To get a variable fidelity data we vary

the grid size for a one-dimensional integration: we generate the low fidelity data using the grid of size 3 and the high fidelity data using the grid of size 1000. We note that if the size of the grid is greater than 3, then the high and low fidelity functions become indistinguishable.

We provide errors in Table 2, which show that the best results are typically obtained using the proposed **MinMinimax** approach. However, there are two drawbacks: sometimes it is impossible to improve the model accuracy using variable fidelity data; or too small sample size is selected making it impossible to construct a reliable regression model. For example, for the **Supernova** dataset the **MinMinimax** method works poorly because it suggests to use the high fidelity sample size equal to four, which is obviously insufficient for the cokriging to work efficiently. That is why we suggest to impose a lower bound for the size of the high fidelity data sample.

6 Conclusions

We prove the minimax interpolation error for the Gaussian process regression in the multivariate case. The obtained results are used to estimate the interpolation error for the regression modeling with the variable fidelity data. This allows us to identify settings in which the accuracy of the regression model can be improved with the variable fidelity data.

Moreover, we estimate the optimal ratio of sizes of the variable fidelity data samples. Using both synthetic and real problems, we demonstrate that this ratio can be used when producing a design of experiments.

However, there is still room for improvement of the proposed approach: it requires an accurate estimate of the correlation coefficient, and it doesn't take into account inaccuracies of estimates of the regression model parameters. Furthermore, in this paper we consider the case of two fidelity levels only, whereas in practice multiple fidelity levels can be accessible.

Acknowledgments

The research, presented in Section 5 of this paper, was supported by the RFBR grants 16-01-00576 A and 16-29-09649 of_m; the research, presented in other sections, was supported solely by the Russian Science Foundation grant (project 14-50-00150).

The authors would like to thank Yu. Golubev for his helpful comments, I. Nazarov and M. Panov for their help with proofreading.

A PROOFS FOR SUBSECTION 2.1

Proof of Theorem 1. It is easy to see that

$$\begin{aligned} \mathbb{E}[f(\mathbf{x}) - \tilde{f}(\mathbf{x})]^2 &= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k} \in \mathbb{Z}^d} K(\mathbf{x} - \mathbf{x}_{\mathbf{k}}) \exp(-2\pi i \boldsymbol{\omega}^T (\mathbf{x}_{\mathbf{k}} - \mathbf{x})) \right|^2 d\boldsymbol{\omega} = \\ &= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k} \in \mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \hat{K}(\mathbf{u}) \exp(-2\pi i \mathbf{u}^T (\mathbf{x} - \mathbf{x}_{\mathbf{k}})) d\mathbf{u} \right) \exp(-2\pi i \boldsymbol{\omega}^T (\mathbf{x}_{\mathbf{k}} - \mathbf{x})) \right|^2 d\boldsymbol{\omega}, \end{aligned}$$

where $\hat{K}(\mathbf{u})$ is the Fourier transform of $K(\mathbf{x})$. As Poisson summation formula suggests:

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \exp(2\pi i \mathbf{k}^T \boldsymbol{\omega}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \delta(\boldsymbol{\omega} + \mathbf{k}),$$

where $\delta(\boldsymbol{\omega})$ is the Dirac delta function, then

$$\begin{aligned}\mathbb{E}[f(\mathbf{x}) - \tilde{f}(\mathbf{x})]^2 &= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k} \in \mathbb{Z}^d} \int_{\mathbb{R}^d} \hat{K}(\mathbf{u}) \exp(2\pi i(\boldsymbol{\omega} - \mathbf{u})^T \mathbf{x}) \delta(\mathbf{u} - \boldsymbol{\omega} + H^{-1}\mathbf{k}) d\mathbf{u} \right|^2 d\boldsymbol{\omega} = \\ &= \int_{\mathbb{R}^d} \left| 1 - \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{K}(\boldsymbol{\omega} - H^{-1}\mathbf{k}) \exp(2\pi i H^{-1} \mathbf{x}^T \mathbf{k}) \right|^2 d\boldsymbol{\omega}.\end{aligned}$$

Taking into account orthogonality of the system of functions $\exp(2\pi i H^{-1} \mathbf{x}^T \mathbf{k})$ on $\mathbf{x} \in [0, h_1] \times \dots \times [0, h_d]$ we integrate the equality to get the interpolation error

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| [1 - \hat{K}(\boldsymbol{\omega})]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

To get $\hat{K}(\boldsymbol{\omega})$ that minimizes the interpolation error we rewrite it as

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} \left| [1 - \hat{K}(\boldsymbol{\omega})]^2 F(\boldsymbol{\omega}) + \hat{K}(\boldsymbol{\omega})^2 \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

To minimize this error we solve this quadratic optimization problem for each $\boldsymbol{\omega}$ and get:

$$\hat{K}(\boldsymbol{\omega}) = \frac{\hat{F}(\boldsymbol{\omega})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}.$$

Then

$$\sigma_H^2(\tilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})} d\boldsymbol{\omega}. \quad (9)$$

□

Proof of Remark 1. It holds that the best approximation has the form

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) f(\mathbf{x}_{\mathbf{k}})$$

for some $\phi(\mathbf{x}, \mathbf{x}')$. As Wiener-Hopf equations for the covariance function $R(\mathbf{x})$ hold, then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}}) = R(\mathbf{x} - \mathbf{x}_{\mathbf{m}}) \quad (10)$$

for all $\mathbf{m} \in \mathbb{Z}^d$. Let us prove that $\phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) = \phi(\mathbf{x} - \mathbf{x}_{\mathbf{k}})$.

Let us consider two sums from (10):

$$\begin{aligned}\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}}) &= R(\mathbf{x} - \mathbf{x}_{\mathbf{m}}), \\ \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x}_{\mathbf{s}}, \mathbf{x}_{\mathbf{k}}) R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}-\mathbf{s}}) &= R(\mathbf{x} - \mathbf{x}_{\mathbf{s}} - \mathbf{x}_{\mathbf{m}-\mathbf{s}}).\end{aligned}$$

As $\mathbf{x}_{\mathbf{m}-\mathbf{s}} = H\mathbf{m} - H\mathbf{s} = \mathbf{x}_{\mathbf{m}} - \mathbf{x}_{\mathbf{s}}$, then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}}) = R(\mathbf{x} - \mathbf{x}_{\mathbf{m}}) = R(\mathbf{x} - \mathbf{x}_{\mathbf{s}} - \mathbf{x}_{\mathbf{m}-\mathbf{s}}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x}_{\mathbf{s}}, \mathbf{x}_{\mathbf{k}}) R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}-\mathbf{s}}).$$

Consequently,

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} [\phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) - \phi(\mathbf{x} - \mathbf{x}_{\mathbf{s}}, \mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{s}})] R(\mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{m}}) = 0.$$

Positive definiteness of the covariance function $R(\mathbf{x})$ implies that

$$\phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) = \phi(\mathbf{x} - \mathbf{x}_{\mathbf{s}}, \mathbf{x}_{\mathbf{k}} - \mathbf{x}_{\mathbf{s}}).$$

For $\mathbf{x}_{\mathbf{s}} = \mathbf{x}_{\mathbf{k}}$ we get

$$\phi(\mathbf{x}, \mathbf{x}_{\mathbf{k}}) = \phi(\mathbf{x} - \mathbf{x}_{\mathbf{k}}, \mathbf{0}) = K(\mathbf{x} - \mathbf{x}_{\mathbf{k}}).$$

Due to Poisson summation formula it holds that

$$\frac{1}{\mu(\Omega_H)} \Phi(\boldsymbol{\omega}) \sum_{\mathbf{k} \in \mathbb{Z}^d} F(\boldsymbol{\omega} - \boldsymbol{\omega}_{\mathbf{k}}) = F(\boldsymbol{\omega}),$$

where $\Phi(\boldsymbol{\omega})$ is the Fourier transform of $\phi(\mathbf{x})$. Then

$$\Phi(\boldsymbol{\omega}) = \frac{\mu(\Omega_H) F(\boldsymbol{\omega})}{\sum_{\mathbf{k} \in \mathbb{Z}^d} F(\boldsymbol{\omega} - \boldsymbol{\omega}_{\mathbf{k}})}.$$

So, optimal interpolation has the form:

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{k} \in \mathbb{Z}^d} K(\mathbf{x} - \mathbf{x}_{\mathbf{k}}) f(\mathbf{x}_{\mathbf{k}}).$$

Also

$$\hat{K}(\boldsymbol{\omega}) = \frac{\Phi(\boldsymbol{\omega})}{\mu(\Omega_H)}.$$

□

Proof of Corollary 1. We get the interpolation error for an exponential covariance function of the form $R(x) = \sqrt{\frac{\pi}{2}} \exp(-\theta|x|)$ for $x \in \mathbb{R}$. The spectral density for this covariance function is $F(\omega) = \frac{\theta}{\theta^2 + \omega^2}$.

We want to evaluate the interpolation error

$$\sigma_h^2(\tilde{f}, F) = \int_{-\infty}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + \frac{k}{h})}{\sum_k F(\omega + \frac{k}{h})} d\omega.$$

It holds that

$$\begin{aligned} \sum_{k=-\infty}^{\infty} F(\omega + \frac{k}{h}) &= \sum_{k=-\infty}^{\infty} \frac{\theta}{(\omega + \frac{k}{h})^2 + \theta^2} = h \sum_{k=-\infty}^{\infty} \frac{h\theta}{(h\omega + k)^2 + h^2\theta^2} = \\ &= \pi h \coth(\pi\theta h) \frac{1}{1 + \sin^2(\pi h\omega)(\coth^2(\pi\theta h) - 1)}. \end{aligned}$$

Then

$$\int_{-\infty}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + \frac{k}{h})}{\sum_k F(\omega + \frac{k}{h})} d\omega = \int_{-\infty}^{\infty} \frac{\theta}{\theta^2 + \omega^2} \left(1 - \frac{\theta}{\theta^2 + \omega^2} \frac{1 + \sin^2(\pi h\omega)(\coth^2(\pi\theta h) - 1)}{\pi h \coth(\pi\theta h)} \right) d\omega.$$

We can integrate three terms inside the integral analytically. Namely,

$$\int_{-\infty}^{\infty} \frac{\theta}{\theta^2 + \omega^2} d\omega = \pi.$$

Also

$$\int_{-\infty}^{\infty} \frac{\theta^2}{(\theta^2 + \omega^2)^2} d\omega = \frac{\pi}{2\theta}.$$

Finally

$$\int_{-\infty}^{\infty} \frac{\theta^2}{(\theta^2 + \omega^2)^2} \sin^2(\pi\omega h) d\omega = -\frac{\pi^2 h}{2} (\cosh(\pi\theta h) - \sinh(\pi\theta h)) \left(\cosh(\pi\theta h) - \left(\frac{1}{\pi\theta h} + 1 \right) \sinh(\pi\theta h) \right).$$

Consequently,

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{\theta}{\theta^2 + \omega^2} \left(1 - \frac{\theta}{\theta^2 + \omega^2} \frac{1 + \sin^2(\pi h \omega) (\coth^2(\pi\theta h) - 1)}{\pi h \coth(\pi\theta h)} \right) d\omega = \\ & \pi - \frac{\pi}{2\pi\theta h \coth(\pi\theta h)} + \\ & + \frac{\pi^2}{2} \exp(-\pi\theta h) \left(\exp(-\pi\theta h) - \frac{1}{\pi\theta h} \sinh(\pi\theta h) \right) \frac{\coth^2(\pi\theta h) - 1}{\coth(\pi\theta h)}. \end{aligned}$$

For $h \rightarrow 0$ we get Taylor series for the obtained interpolation error:

$$\sigma_h^2(\tilde{f}, F) = \frac{2\pi^2}{3} \theta h + O((\theta h)^2).$$

□

Proof of Corollary 2. Note, that the interpolation error has the form

$$\sigma_h^2(\tilde{f}, F) = \int_{-\infty}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + \frac{k}{h})}{\sum_s F(\omega + \frac{s}{h})} d\omega.$$

We get lower and upper bounds for this expression. We denote $v = \frac{1}{h}$.

We get upper bound for the interpolation error by splitting integration region $(-\infty, \infty)$ to three intervals $(-\infty, -v/2]$, $(-v/2, v/2]$, $(v/2, +\infty)$ and obtaining an upper bound for each of them.

Note that

$$0 \leq \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} \leq 1.$$

Consequently, using Chernov type bounds [11] we get

$$\begin{aligned} & \int_{v/2}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \leq \int_{v/2}^{\infty} F(\omega) d\omega = \\ & = \int_{v/2}^{\infty} \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right) d\omega \leq \sqrt{2} \exp\left(-\frac{v^2}{8\theta}\right). \end{aligned} \tag{11}$$

In a similar way get an upper bound for the interval $(-\infty, -v/2)$.

Now we get an estimate for the interval $(-v/2, v/2)$. We start with an upper bound and a lower bound for series $\sum_{s \neq 0} F(\omega + sv)$. Spectral density for squared exponential covariance function decreases at $[0, +\infty)$ with respect to ω . Thus,

$$\int_{\Delta+u}^{+\infty} F(x) dx \leq \sum_{s=1}^{\infty} \Delta F(\Delta s + u) \leq \Delta F(s + u) + \int_{\Delta+u}^{+\infty} F(x) dx.$$

Using [1], Formula 7.1.13, we get for ω such that $|\omega| \leq \frac{v}{2}$:

$$\frac{4\sqrt{\theta}}{v + \sqrt{v^2 + 16\theta}} \exp\left(-\frac{v^2}{8\theta}\right) \leq \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

And

$$v \sum_{k \in \mathbb{Z}^+} F(\omega + kv) \leq vF(\omega + v) + \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \frac{v}{\sqrt{\theta}} \exp\left(-\frac{v^2}{8\theta}\right) + \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

Now we are ready to get an upper bound for the integral over the interval $(-v/2, v/2)$ for big enough v :

$$\begin{aligned} & \int_{-v/2}^{v/2} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \leq \\ & \leq \int_{-v/2}^{v/2} F(\omega) \frac{F(\omega + v) + F(\omega - v) + \frac{4\sqrt{\theta}}{v(v + \sqrt{v^2 + \frac{32}{\pi}\theta})} \exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega) + F(\omega + v) + F(\omega - v) + \frac{4\sqrt{\theta}}{v(v + \sqrt{v^2 + \frac{32}{\pi}\theta})} \exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq \\ & \leq \int_{-v/2}^{v/2} F(\omega) \frac{F(\omega + v) + F(\omega - v)}{F(\omega) + F(\omega + v) + F(\omega - v)} d\omega + \int_{-v/2}^{v/2} F(\omega) \frac{\frac{4\sqrt{\theta}}{v(v + \sqrt{v^2 + \frac{32}{\pi}\theta})} \exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega) + \frac{4\sqrt{\theta}}{v(v + \sqrt{v^2 + \frac{32}{\pi}\theta})} \exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq \\ & \leq \int_{-v/2}^{v/2} F(\omega + v) + F(\omega - v) d\omega + \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right) \leq \\ & \leq \frac{12\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right) \leq \frac{7\sqrt{\theta}}{v} \exp\left(-\frac{v^2}{8\theta}\right). \end{aligned}$$

It holds that

$$\frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} \geq \frac{F(\omega + v) + F(\omega - v)}{F(\omega) + F(\omega + v) + F(\omega - v)}.$$

For ω such that $|\omega| \leq \frac{v}{2}$ we get that:

$$1 + \frac{F(\omega + v)}{F(\omega)} + \frac{F(\omega - v)}{F(\omega)} \leq 3.$$

Then for sufficiently large v the following lower bound holds:

$$\begin{aligned} & \int_{-\infty}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \geq \int_{-v/2}^{v/2} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \geq \\ & \geq \int_{-v/2}^{v/2} F(\omega) \frac{F(\omega + v) + F(\omega - v)}{F(\omega) + F(\omega + v) + F(\omega - v)} d\omega \geq \int_{-v/2}^{v/2} \frac{F(\omega + v) + F(\omega - v)}{3} d\omega = \\ & = \frac{2}{3} \int_{-v/2}^{v/2} F(\omega) d\omega \geq \frac{2}{3} \left(\frac{4\sqrt{\theta}}{v + \sqrt{v^2 + 16\theta}} \exp\left(-\frac{v^2}{8\theta}\right) - \frac{4\sqrt{\theta}}{3v + \sqrt{9v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{9v^2}{8\theta}\right) \right) \geq \\ & \geq \frac{4}{3} \frac{\sqrt{\theta}}{v} \exp\left(-\frac{v^2}{8\theta}\right). \end{aligned}$$

□

B PROOFS FOR SUBSECTION 2.2

We need the following lemma to complete the proof of the main result

Lemma 1. *Let $c \geq 0$ and $\omega \geq 0$ be such that $c^2 + \omega^2 \leq 1$, $c^2 + (1 - \omega^2) \leq 1$. Then*

$$\left(1 - \sqrt{c^2 + \omega^2}\right)^2 + \left(1 - \sqrt{c^2 + (1 - \omega^2)}\right)^2 \leq \left(1 - \sqrt{c^2}\right)^2 = (1 - c)^2. \quad (12)$$

Proof. We start with a scheme of the proof. We prove that for ω , that maximizes left hand side of the inequality (12), this inequality holds. To prove this we show that for admissible $\omega \in [1 - \sqrt{1 - c^2}, \frac{1}{2}]$ derivative of the left hand side with respect to ω is smaller than zero for all admissible c , so $\omega = 1 - \sqrt{1 - c^2}$ provides maximum of the left hand side, and for such ω inequality holds.

Partial derivative of the left hand side with respect to ω is equal to

$$\begin{aligned} g(\omega, c) &= \frac{\partial}{\partial \omega} \left(\left(1 - \sqrt{c^2 + \omega^2}\right)^2 + \left(1 - \sqrt{c^2 + (1 - \omega)^2}\right)^2 \right) = \\ &= -2 \frac{(1 - \sqrt{c^2 + \omega^2}) \omega}{\sqrt{c^2 + \omega^2}} + 2 \frac{\left(1 - \sqrt{c^2 + (1 - \omega)^2}\right) (1 - \omega)}{\sqrt{c^2 + (1 - \omega)^2}} = \\ &= -2 \left(\frac{1}{\sqrt{c^2 + \omega^2}} - 1 \right) \omega + 2 \left(\frac{1}{\sqrt{c^2 + (1 - \omega)^2}} - 1 \right) (1 - \omega). \end{aligned}$$

If $\omega = \frac{1}{2}$, then the partial derivative is zero. We show that for such $\omega < \frac{1}{2}$ that $c^2 + \omega^2 < 1$, $c^2 + (1 - \omega)^2 < 1$, it holds that $g(\omega, c) < 0$. This fact means that the initial function decreases for $\omega \in [1 - \sqrt{1 - c^2}, \frac{1}{2}]$.

We start with maximization of $g(\omega, c)$ with respect to c . The function $g(\omega, c)$ attains maximum at the edge of admissibility region or in a local optimum with respect to c . To find local optima we search for c , such that the partial derivative $g(\omega, c)$ with respect to c is equal to zero:

$$\frac{c(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{3}{2}}} - \frac{c\omega}{(\omega^2 + c^2)^{\frac{3}{2}}} = 0.$$

Consequently,

$$\frac{1 - \omega}{\omega} = \frac{((1 - \omega)^2 + c^2)^{\frac{3}{2}}}{(\omega^2 + c^2)^{\frac{3}{2}}}. \quad (13)$$

So,

$$c^2 = \omega^{\frac{2}{3}}(1 - \omega)^{\frac{2}{3}}(\omega^{\frac{2}{3}} + (1 - \omega)^{\frac{2}{3}}).$$

We show that this is a local maximum. Namely, we prove that the second partial derivative of $g(\omega, c)$ with respect to c is smaller than 0:

$$-\frac{(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{3}{2}}} + \frac{\omega}{(\omega^2 + c^2)^{\frac{3}{2}}} + \frac{3c^2(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{5}{2}}} - \frac{3c^2\omega}{(\omega^2 + c^2)^{\frac{5}{2}}} \leq 0.$$

Or:

$$\frac{\omega}{(\omega^2 + c^2)^{\frac{5}{2}}} ((\omega^2 + c^2) - 3c^2) - \frac{(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{5}{2}}} (((1 - \omega)^2 + c^2) - 3c^2) \leq 0.$$

In a local optimum (13) holds, and we can rewrite inequality as:

$$\frac{(1 - \omega)}{(\omega^2 + c^2)((1 - \omega)^2 + c^2)^{\frac{3}{2}}} (\omega^2 - 2c^2) - \frac{(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{5}{2}}} ((1 - \omega)^2 - 2c^2) \leq 0.$$

Then,

$$\frac{(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{5}{2}}(\omega^2 + c^2)} (((1 - \omega)^2 + c^2)(\omega^2 - 2c^2) - (\omega^2 + c^2)((1 - \omega)^2 - 2c^2)) \leq 0.$$

Due to constraints on values of ω this is equivalent to:

$$((1 - \omega)^2 + c^2)(\omega^2 - 2c^2) - (\omega^2 + c^2)((1 - \omega)^2 - 2c^2) \leq 0,$$

or

$$2c^2\omega^2 - c^2(1-\omega)^2 - 2c^2(1-\omega)^2 + c^2\omega^2 \leq 0.$$

This inequality holds, as $\omega \leq \frac{1}{2}$ and $(1-\omega)^2 \geq \omega^2$.

So, the extremum is a local minimum, and the function attains maximum values at the edges of the admissibility region. Namely, $c^2 = 1 - (1-\omega)^2$ or $c^2 = 0$ provides maximum values.

For such values of c the derivative is smaller than zero. Using $c^2 = 1 - (1-\omega)^2$ we get —

$$-2 \left(\frac{1}{\sqrt{1 - (1-\omega)^2 + \omega^2}} - 1 \right) \omega \leq 0.$$

In a similar way for $c^2 = 0$

$$-2(1-\omega) + 2\omega = 4\omega - 2 \leq 0.$$

Consequently, the target function decreases with respect to ω at $[1 - \sqrt{1 - c^2}, \frac{1}{2}]$, and $\omega = \frac{1}{2}$ provides a local minimum. So, the local maximum for left hand side is at $\omega = 1 - \sqrt{1 - c^2}$. It is easy to see that in this case the left side of (12) is not larger than the right side. \square

Let us now prove the main theorem.

Proof of Theorem 2. We provide upper and lower bounds for $R^H(L, \boldsymbol{\lambda})$ that are equal to $\frac{L}{2\pi^2} \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i} \right)^2$. We start with a lower bound, and then continue with an upper bound.

We consider a functional

$$\Phi(F, \hat{K}) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[(1 - \hat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] d\boldsymbol{\omega},$$

that is equal to the interpolation error $\sigma_H^2(\tilde{f}, F)$ for

$$\tilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}'),$$

such that $\hat{K}(\boldsymbol{\omega})$ is the Fourier transform of $K(\mathbf{x})$.

The functional is linear in $F(\boldsymbol{\omega})$ and quadratic in $\hat{K}(\boldsymbol{\omega})$, and we search for a saddle point of the functional $R^H(L, \boldsymbol{\lambda})$ such that:

$$R^H(L, \boldsymbol{\lambda}) = \inf_{\tilde{f}} \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \sigma_H^2(\tilde{f}, F) = \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \inf_{\tilde{f}} \sigma_H^2(\tilde{f}, F).$$

It holds that (9)

$$\min_{\hat{K}} \Phi(F, \hat{K}) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} F(\boldsymbol{\omega} + \mathbf{x})}{\sum_{\mathbf{x} \in D_{H-1}} F(\boldsymbol{\omega} + \mathbf{x})} d\boldsymbol{\omega}.$$

Let us consider a class of spectral densities $F_\varepsilon(\boldsymbol{\omega})$:

$$F_\varepsilon(\boldsymbol{\omega}) = \begin{cases} \frac{A_\varepsilon}{(2\varepsilon)^d}, & \exists \mathbf{s} \in U_h : \|\boldsymbol{\omega} - \mathbf{s}\|_\infty \leq \varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

here $U_h = \left\{ \left(0, 0, \dots, \frac{1}{2h_j}, \dots, 0 \right), \left(0, 0, \dots, -\frac{1}{2h_j}, \dots, 0 \right) \right\}$, and an index j is such that

$$j = \arg \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i} \right)^2.$$

Due to (3)

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^d \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \leq L,$$

and for $\varepsilon \rightarrow 0$

$$A_\varepsilon \rightarrow \frac{L}{2\pi^2} \left(\frac{h_j}{\lambda_j} \right)^2.$$

Really, for $\varepsilon \rightarrow 0$:

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^d \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \rightarrow 2(2\pi)^2 \frac{A_\varepsilon}{(2\varepsilon)^d} (2\varepsilon)^d \left(\frac{\lambda_j}{h_j} \right)^2 = 2A_\varepsilon \left(\frac{\pi \lambda_j}{h_j} \right)^2 = L.$$

Now for $\varepsilon \rightarrow 0$ it holds that

$$\min_{\hat{K}} \Phi(F_\varepsilon, \hat{K}) \rightarrow 2 \frac{1}{2} \frac{A_\varepsilon}{(2\varepsilon)^d} (2\varepsilon)^d = A_\varepsilon = \frac{L}{2\pi^2} \left(\frac{h_j}{\lambda_j} \right)^2.$$

Consequently, we get a lower bound that equals $\frac{L}{2\pi^2} \left(\frac{h_j}{w_j} \right)^2$. Now we continue with a proof of the upper bound.

For any $\hat{K}(\boldsymbol{\omega})$ it holds that

$$\begin{aligned} R^H(L, \boldsymbol{\lambda}) &\leq \max_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \Phi(F, \hat{K}) \leq \\ &\leq L \left(\frac{1}{2\pi} \right)^2 \max_{\boldsymbol{\omega}} \left\{ \frac{1}{\sum_{i=1}^d \lambda_i^2 \omega_i^2} \left[(1 - \hat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] \right\}. \end{aligned}$$

Now let us consider

$$\hat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \|\boldsymbol{\omega}\|, & \|\boldsymbol{\omega}\|^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Now we prove that for such $\hat{K}(\boldsymbol{\omega})$ it holds that

$$\left[(1 - \hat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] \leq 2\|\boldsymbol{\omega}\|^2. \quad (14)$$

It holds that $(1 - \hat{K}(\boldsymbol{\omega}))^2 \leq \|\boldsymbol{\omega}\|^2$. Now let us prove that

$$\sum_{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \leq \|\boldsymbol{\omega}\|^2. \quad (15)$$

We use mathematical induction by d for $\boldsymbol{\omega}$ such that $\|\boldsymbol{\omega}\|_\infty < 1$. We prove that for $\|\boldsymbol{\omega}\|_\infty < 1$ and $c^2 \geq 0$:

$$\sum_{\substack{\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}, \\ \|\boldsymbol{\omega} + \mathbf{x}\|^2 + c^2 \leq 1}} \left(1 - \sqrt{c^2 + \sum_{i=1}^d (\omega_i + x_i)^2} \right)^2 \leq \sum_{\substack{i \in \{1, \dots, d\}, \\ c^2 + (1 - \omega_i)^2 \leq 1}} \left(1 - \sqrt{c^2 + (1 - \omega_i)^2} \right)^2.$$

For $d = 1$ the induction statement is trivial, as the right hand side and the left hand side coincide. Suppose that for $(d - 1)$ the induction statement holds. Now let us prove that the induction statement holds for d .

For $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_d)$ such that $\|\boldsymbol{\omega}\|_\infty < 1$, i -th component of the vector $\boldsymbol{\omega} + \mathbf{x}$, $\mathbf{x} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ such that $\|\boldsymbol{\omega} + \mathbf{x}\| \leq 1$ is either ω_i or $(1 - \omega_i)$. Consequently, all such $\boldsymbol{\omega} + \mathbf{x}$ has the form $\mathbf{s}\boldsymbol{\omega} + (\mathbf{1} - \mathbf{s})(\mathbf{1} - \boldsymbol{\omega})$, where \mathbf{s} is a vector with all components of it belong to $\{0, 1\}$.

It holds that $(1 - \sqrt{c^2 + (1 - \omega_1)^2 + \omega_2^2 + \dots + \omega_d^2})^2 \leq (1 - \sqrt{c^2 + (1 - \omega_1)^2})^2$, if $c^2 + (1 - \omega_1)^2 + \omega_2^2 + \dots + \omega_d^2 \leq 1$.

Now let us consider all terms of the form $(1 - \sqrt{c^2 + (1 - \omega_1)^2 + \dots})^2$ for which there exists $j \neq 1$, such that $(1 - \omega_j)^2$ is in the sum inside the squared root. Due to the induction statement sum of these terms is bounded by:

$$\sum_{\substack{i \in \{2, \dots, d\}, \\ c^2 + (1 - \omega_1)^2 + (1 - \omega_i)^2 \leq 1}} (1 - \sqrt{c^2 + (1 - \omega_1)^2 + (1 - \omega_i)^2})^2.$$

In the same way we prove that the sum of terms $(1 - \sqrt{c^2 + \omega_1^2 + \dots})^2$ with a term $(1 - \omega_j)^2$ inside the root is upper bounded by:

$$\sum_{\substack{i \in \{2, \dots, d\}, \\ c^2 + \omega_1^2 + (1 - \omega_i)^2 \leq 1}} (1 - \sqrt{c^2 + \omega_1^2 + (1 - \omega_i)^2})^2.$$

Using Lemma 1 for a pair of terms $(1 - \sqrt{c^2 + (1 - \omega_1)^2 + (1 - \omega_i)^2})^2 + (1 - \sqrt{c^2 + \omega_1^2 + (1 - \omega_i)^2})^2$ we get:

$$\begin{aligned} & (1 - \sqrt{c^2 + (1 - \omega_1)^2 + (1 - \omega_i)^2})^2 + (1 - \sqrt{c^2 + \omega_1^2 + (1 - \omega_i)^2})^2 \leq \\ & \leq (1 - \sqrt{c^2 + (1 - \omega_i)^2})^2. \end{aligned}$$

This upper bound also holds if there are no or only one term for i -th index. Consequently, the induction statement holds: the target sum is bounded by $\sum_{i=1}^d (1 - \sqrt{c^2 + (1 - \omega_i)^2})^2$.

Using $c^2 = 0$ we get (15).

Now let us consider the case $\|\boldsymbol{\omega}\|_\infty \geq 1$. We look at the case $\boldsymbol{\omega} = \{\hat{\omega}_1 + 1, \omega_2, \dots, \omega_d\}$, moreover $\|(\hat{\omega}_1, \omega_2, \dots, \omega_d)\|_\infty < 1$, and $\hat{\omega}_1 \geq 0$, $\omega_i \geq 0$, $i = \overline{2, d}$. Then $\|\boldsymbol{\omega}\|^2 = 1 + 2\hat{\omega}_1 + \hat{\omega}_1^2 + \sum_{i=2}^d \omega_i^2$. For vector $(\hat{\omega}_1, \omega_2, \dots, \omega_d)$ we have the induction statement above (15). For the initial vector $\boldsymbol{\omega}$ we have an additional term $\hat{K}^2((\hat{\omega}_1, \omega_2, \dots, \omega_d))$ if Euclidian norm of such a vector is below or equal 1 — but this new term is smaller or equal to 1, as otherwise this term is not in the sum. So, the target estimate for $\boldsymbol{\omega}$ holds. Other cases for $\|\boldsymbol{\omega}\|_\infty > 1$ are similar. Consequently for all $\boldsymbol{\omega}$ the estimate (14) holds.

It holds that

$$\max_{\boldsymbol{\omega}} \frac{\sum_{i=1}^d \omega_i^2}{\sum_{i=1}^d \left(\frac{\lambda_i}{h_i}\right)^2 \omega_i^2} = \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Consequently, an upper bound for minimax interpolation error holds

$$R^H(L, \boldsymbol{\lambda}) \leq \frac{L}{2\pi^2} \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

The upper bound coincides with the lower bound. The theorem holds. \square

C PROOFS FOR SUBSECTION 3.2

Proof of Theorem 3. For convenience we redefine all points that belong to D_H as $D_H = \{\mathbf{x}_i\}$ and all points that belong to $D_{\frac{H}{m}}$ as $D_{\frac{H}{m}} = \{\tilde{\mathbf{x}}_j\}$. Then for Gaussian process regression the best unbiased estimator is linear in known values:

$$\tilde{u}(\mathbf{x}) = \sum_i k_i u(\mathbf{x}_i) + \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j).$$

for some k_i, \tilde{k}_j . Our problem is then to find coefficients k_i, \tilde{k}_j that minimize $\mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2$. Using independence of random processes $f(\mathbf{x})$ and $g(\mathbf{x})$ we get:

$$\begin{aligned} \mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2 &= \mathbb{E} \left[\rho f(\mathbf{x}) + g(\mathbf{x}) - \sum_i k_i(\rho f(\mathbf{x}_i) + g(\mathbf{x}_i)) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j) \right]^2 = \\ &= \mathbb{E} \left[\rho f(\mathbf{x}) - \sum_i \rho k_i f(\mathbf{x}_i) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j) \right]^2 + \mathbb{E} \left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i) \right]^2. \end{aligned}$$

For each i there exists an index j such $\mathbf{x}_i = \tilde{\mathbf{x}}_j$. Denote

$$\tilde{k}'_j = \begin{cases} \frac{1}{\rho} \tilde{k}_j, & \forall i, \tilde{\mathbf{x}}_j \neq \mathbf{x}_i, \\ \frac{1}{\rho} \tilde{k}_j + k_i, & \exists i, \tilde{\mathbf{x}}_j = \mathbf{x}_i. \end{cases}$$

There is a one-to-one correspondence between $(\{k_i\}, \{\tilde{k}_j\})$ and $(\{k_i\}, \{\tilde{k}'_j\})$, so minimization of $\mathbb{E}(u(\mathbf{x}) - \tilde{u}(\mathbf{x}))^2$ with respect to k_i, \tilde{k}_j is equivalent to minimization of this function with respect to k_i, \tilde{k}'_j . Then

$$\begin{aligned} &\mathbb{E} \left[\rho f(\mathbf{x}) - \sum_i k_i \rho f(\mathbf{x}_i) - \sum_j \tilde{k}_j f(\tilde{\mathbf{x}}_j) \right]^2 + \mathbb{E} \left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i) \right]^2 = \\ &= \rho^2 \mathbb{E} \left[f(\mathbf{x}) - \sum_j \tilde{k}'_j f(\tilde{\mathbf{x}}_j) \right]^2 + \mathbb{E} \left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i) \right]^2. \end{aligned}$$

For terms $\mathbb{E} \left[f(\mathbf{x}) - \sum_j \tilde{k}'_j f(\tilde{\mathbf{x}}_j) \right]^2$ and $\mathbb{E} [g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)]^2$ minimization problems are equivalent to that of single fidelity data — and the first term contains only coefficients \tilde{k}'_j , the second term contains only coefficients k_i .

For k_i and \tilde{k}'_j that minimize interpolation error at point for the single fidelity scenario it holds that $\tilde{k}'_j = K_f(\mathbf{x} - \mathbf{x}_j)$, $k_i = K_g(\mathbf{x} - \mathbf{x}_i)$ for some symmetric kernels $K_f(\mathbf{x} - \mathbf{x}_j)$, $K_g(\mathbf{x} - \mathbf{x}_i)$.

Now we continue proof for $f(\mathbf{x})$ and $g(\mathbf{x})$ in a way similar to the single fidelity case. For $\mathbb{E} [g(\mathbf{x}) - \sum_i K_g(\mathbf{x} - \mathbf{x}_i) g(\mathbf{x}_i)]^2$ it holds that

$$\begin{aligned} &\frac{1}{|H|} \int_{x_i \in [0, h_i], i=1, d} \mathbb{E} \left[g(\mathbf{x}) - \sum_i K_g(\mathbf{x} - \mathbf{x}_i) g(\mathbf{x}_i) \right]^2 d\mathbf{x} = \\ &= \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \left[[1 - \hat{K}_g(\boldsymbol{\omega})]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_g^2(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right] d\boldsymbol{\omega} \end{aligned}$$

In a similar way we get for the interval $[0, \frac{h_1}{m}] \dots [0, \frac{h_d}{m}]$ for $\mathbb{E} \left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j) \right]^2$:

$$\begin{aligned} &\frac{m^d}{|H|} \int_{x_i \in [0, \frac{h_i}{m}], i=1, d} \mathbb{E} \left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j) \right]^2 d\mathbf{x} = \\ &= \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[[1 - \hat{K}_f(\boldsymbol{\omega})]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right] d\boldsymbol{\omega}. \end{aligned}$$

Consequently,

$$\begin{aligned} & \frac{1}{|H|} \int_{x_i \in [0, h_i], i=1, d} \mathbb{E} \left[f(\mathbf{x}) - \sum_j K_f(\mathbf{x} - \tilde{\mathbf{x}}_j) f(\tilde{\mathbf{x}}_j) \right]^2 d\mathbf{x} = \\ & = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_f(\boldsymbol{\omega}) \right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2(\boldsymbol{\omega} + mH^{-1}\mathbf{k}) \right] d\boldsymbol{\omega}. \end{aligned}$$

So, the target interpolation error (5) has the form:

$$\begin{aligned} \sigma_{H,m}^2(\tilde{u}, F, G, \rho) &= \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_g(\boldsymbol{\omega}) \right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_g^2(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right] d\boldsymbol{\omega} + \\ &+ \rho^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \hat{K}_f(\boldsymbol{\omega}) \right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \hat{K}_f^2(\boldsymbol{\omega} + mH^{-1}\mathbf{k}) \right] d\boldsymbol{\omega}. \end{aligned}$$

Finally,

$$\sigma_{H,m}^2(\tilde{u}, F, G, \rho) = \sigma_H^2(\tilde{g}, G) + \rho^2 \sigma_{\frac{H}{m}}^2(\tilde{f}, F).$$

□

D PROOFS FOR SECTION 4

In this section we provide a proof of Theorem 5.

Proof of Theorem 5. Minimax interpolation error has the form:

$$R_2 = \frac{L_g}{2} \frac{1}{\pi^2} \left(\frac{c + (m^*)^d}{\Lambda} \right)^{\frac{2}{d}} + \rho^2 \frac{L_f}{2} \frac{1}{\pi^2} \left(\frac{c + (m^*)^d}{(m^*)^d \Lambda} \right)^{\frac{2}{d}}.$$

Denote $\delta = (m^*)^d$. Then we need to minimize with respect to a the following expression

$$\frac{L_g}{2} (c + \delta)^{\frac{2}{d}} + \rho^2 \frac{L_f}{2} \left(\frac{c + \delta}{\delta} \right)^{\frac{2}{d}}.$$

Partial derivative with respect to δ should equal 0:

$$\frac{L_g}{2} (c + \delta)^{\frac{2}{d}-1} \frac{2}{d} + \rho^2 \frac{L_f}{2} \left(\frac{c + \delta}{\delta} \right)^{\frac{2}{d}-1} \frac{2}{d} \frac{-c}{\delta^2} = 0.$$

Consequently

$$\frac{L_g}{2} + \rho^2 \frac{L_f}{2} \delta^{1-\frac{2}{d}} \frac{-c}{\delta^2} = 0.$$

So

$$L_g = L_f \frac{\rho^2 c}{\delta^{1+\frac{2}{d}}}.$$

Finally,

$$\delta = \left(c \rho^2 \frac{L_f}{L_g} \right)^{\frac{d}{d+2}},$$

also we get that

$$m^* = \sqrt[d+2]{c \rho^2 \frac{L_f}{L_g}}.$$

□

References

- [1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1964.
- [2] N. Alexandrov, R. Lewis, C. Gumbert, L. Green, and P. Newman. Optimization with variable-fidelity models applied to wing design. Technical report, DTIC Document, 1999.
- [3] J. Bandler, Q. Cheng, S. Dakrouy, A. Mohamed, M. Bakr, K. Madsen, and J. Søndergaard. Space mapping: the state of the art. *Microwave Theory and Techniques, IEEE Transactions on*, 52(1):337–361, 2004.
- [4] A. Bernstein, E. Burnaev, S. Chernova, F. Zhu, and N. Qin. Comparison of three geometric parameterization methods and their effect on aerodynamic optimization. In *Proceedings of International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems, Eurogen*, pages 14–16, 2011.
- [5] M. Bevilacqua, A. Fassò, C. Gaetan, E. Porcu, and D. Velandia. Covariance tapering for multivariate Gaussian random fields estimation. *Statistical Methods & Applications*, pages 1–17, 2015.
- [6] A. Bhattacharya, D. Pati, and D. Dunson. Anisotropic function estimation using multi-bandwidth Gaussian processes. *Annals of statistics*, 42(1):352, 2014.
- [7] R.K. Bock, A. Chilingarian, M. Gaug, F. Hakl, Th Hengstebeck, M Jiřina, J Klaschka, E Kotrč, P Savický, S Towers, et al. Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 516(2):511–528, 2004.
- [8] E. Burnaev and M. Panov. Adaptive design of experiments based on Gaussian processes. In *Statistical Learning and Data Sciences*, pages 116–125. Springer, 2015.
- [9] E. Burnaev, M. Panov, and A. Zaytsev. Regression on the basis of nonstationary Gaussian processes with Bayesian regularization. *Journal of Communications Technology and Electronics*, 61(6):661–671, 2016.
- [10] E. Burnaev and A. Zaytsev. Surrogate modeling of multifidelity data for large samples. *Journal of Communications Technology and Electronics*, 60(12):1348–1355, 2015.
- [11] S.-H. Chang, P.C. Cosman, and L.B. Milstein. Chernoff-type bounds for the Gaussian error function. *Communications, IEEE Transactions on*, 59(11):2939–2944, 2011.
- [12] A.I.J. Forrester, A. Sóbester, and A.J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2088):3251–3269, 2007.
- [13] G.K. Golubev and E.A. Krymova. On interpolation of smooth processes and functions. *Problems of Information Transmission*, 49(2):127–148, 2013.
- [14] Z. Han, S. Görtz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189, 2013.
- [15] I.A. Ibragimov and Y.A. Rozanov. *Gaussian random processes*, volume 9. Springer Science & Business Media, 2012.
- [16] K. Kandasamy, G. Dasarathy, J. Oliva, J. Schneider, and B. Póczos. Gaussian process optimisation with multi-fidelity evaluations. In *Accepted to NIPS*, 2016.

- [17] M.C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.
- [18] A.N. Kolmogorov. Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 5(1):3–14, 1941.
- [19] S. Koziel, A. Bekasiewicz, I. Couckuyt, and T. Dhaene. Efficient multi-objective simulation-driven antenna design using co-kriging. *Antennas and Propagation, IEEE Transactions on*, 62(11):5900–5905, 2014.
- [20] S. Kucherenko et al. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017, 2009.
- [21] Y. Kuya, K. Takeda, X. Zhang, and A. Forrester. Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA journal*, 49(2):289–298, 2011.
- [22] L. Le Gratiet and C. Cannamela. Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3):418–427, 2015.
- [23] L. Le Gratiet and J. Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 2014.
- [24] S. Pan and Q. Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359, 2010.
- [25] F. Pascual and H. Zhang. Estimation of linear correlation coefficient of two correlated spatial processes. *Sankhyā: The Indian Journal of Statistics*, pages 307–325, 2006.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [27] Pritam Ranjan, Wilson Lu, Derek Bingham, Shane Reese, Brian J Williams, Ch. Chou, F. Doss, M. Grosskopf, and J.P. Holloway. Follow-up experimental designs for computer models and physical processes. *Journal of Statistical Theory and Practice*, 5(1):119–136, 2011.
- [28] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006.
- [29] T. Simpson, V. Toropov, V. Balabanov, and F. Viana. Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, volume 5, pages 10–12, 2008.
- [30] M. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.
- [31] T. Suzuki. PAC-Bayesian bound for gaussian process regression and multiple kernel additive model. In *Proceedings of the 25th annual conference on Computational Learning Theory*, volume 22, pages 39–60, 2012.
- [32] A. van der Vaart and J. van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. *The Annals of Statistics*, pages 1435–1463, 2008.
- [33] D. Velandia, F. Bachoc, M. Bevilacqua, X. Gendre, and J. Loubes. Maximum likelihood estimation for a bivariate Gaussian process under fixed domain asymptotics. *arXiv preprint arXiv:1603.09059*, 2016.

- [34] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press Cambridge, MA, 1949.
- [35] W. Xu, T. Tran, R. Srivastava, and A. Journel. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *SPE annual technical conference and exhibition*. Society of Petroleum Engineers, 1992.
- [36] A. Zaytsev. Variable fidelity regression using low fidelity function blackbox and sparsification. In *Conformal and Probabilistic Prediction with Applications*, pages 147–164. Springer, 2016.
- [37] A.A. Zaytsev, E.V. Burnaev, and V.G. Spokoiny. Properties of the bayesian parameter estimation of a regression based on Gaussian processes. *Journal of Mathematical Sciences*, 203(6):789–798, 2014.
- [38] H. Zhang, W. Cai, et al. When doesnt cokriging outperform kriging? *Statistical Science*, 30(2):176–180, 2015.
- [39] Y. Zhang, J. Duchi, and M. Wainwright. Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *Journal of Machine Learning Research*, 16:3299–3340, 2015.